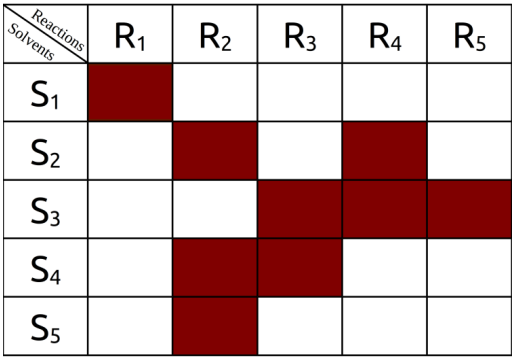
T. N. Akhmetshin [1]
R. I. Nugmanov [1]
T. I. Madzhidov [1]
A. Varnek [1,2]

DEVELOPMENT OF A PREDICTION' PERFORMANCE METRIC FOR CHEMICAL REACTIONS' MODELS

[1] A.M. Butlerov Insitute of Chemistry, Kazan Federal University, Russia
[2] University of Strasbourg, Strasbourg, France

tagirshin@gmail.com

Naive cross-validation is one of procedure for evaluation of predictive ability of chemical reactions, that has the best quality. However, it has a big problem, namely over-estimation of predictive ability. In this work a new procedure for evaluation of predictive ability of chemical reactions is proposed. For correct reaction describing it is necessary to know their conditions (solvent, temperature, etc.), so it makes sense if we estimate the predictive ability of new structures, as well as predictive ability of new conditions. In this case, for evaluation of predictive possibility of new structures we have taken the approach of reactions separating based on their structure during cross-validation. Thus, different reactions with their conditions are only in one fold. From reactions' conditions we evaluated prediction ability in a new solvent, since reasonable well extrapolation of temperature. Furthermore, the number of solvents under one hundred, so it makes sense to use leave-one-out cross validation. Consequently, cooperative approach of cross-validation of structure and solvent predictive ability provides an opportunity to estimate upper and lower limits of models' accuracy.



The research was supported by Russian Science Foundation, grant 14-43-00024.